UNITED STATES DEPARTMENT OF COMMERCE
Bureau of the Census
Washington, DC 20233-0001

OCT 26 2000

MASTER FILE

DSSD CENSUS 2000 PROCEDURES & OPERATIONS MEMORANDUM #0-6R(Revised)

MEMORANDUM FOR Michael J. Longini
Chief, Decennial Systems and Contracts Management Office

From: Howard Hogan
Chief, Decennial Statistical Studies Division

Subject: Specification for the Census 2000 Unduplication of the Service-Based Enumeration Universe

Prepared by: Tracey A. McNally
Decennial Statistical Studies Division

The Service-Based Enumeration (SBE) was conducted over a three day period in March 2000. Emergency and transitional shelters were enumerated on Monday, March 27, soup kitchens and regularly scheduled mobile food vans were enumerated on Tuesday, March 28, and on Wednesday, March 29, we enumerated targeted nonsheltered outdoor locations. This time frame allows for people to be enumerated more than once during this operation. For example, if someone uses a shelter on March 27, and receives services at one or more soup kitchens and/or mobile food vans on March 28, they may be enumerated at each of these services.

This memorandum provides the specifications for unduplicating people without a usual residence who were enumerated more than once during the SBE operation in the Census 2000.

This memorandum also provides specifications for unduplicating people who completed a Be Counted Form (BCF) and marked the "No Address on April 1, 2000" box on that form. These people will be proportionally allocated to shelters and soup kitchens after unduplication within the SBE universe.

Please direct questions concerning these specifications to Tracey McNally, 301-457-4302.

I. UNDUPLICATION SPECIFICATIONS

A. SBE Universe

Data capture records must contain at least two 100% person characteristics to be included in the SBE universe. These include the following records:

1. **Shelter-enumerated persons**
   a. GQ Type Codes: 701, 702, and 703
   b. Persons enumerated on the Individual Census Report (Short form D-20A and long form D-20B)

2. **Soup kitchen and regularly scheduled mobile food van enumerated persons**
   a. GQ Type Codes: 704 and 705
   b. Persons enumerated on the Individual Census Questionnaire (Short form D-15A and long forms D-15B)

3. **Targeted Non-Sheltered Outdoor location enumerated persons**
   a. GQ Type Code: 706
   b. Persons enumerated on the Individual Census Report (Short form D-20A)

4. **Be Counted Form enumerated persons** (Form D-10) meeting the following criteria:
   a. First and last name (combined fields must contain at least 3 alphabetic characters)
   b. The "No address on April 1, 2000" box is marked and a city and/or county and state is provided; *or* the address is a searchable Ususal Home Elsewhere (UHE) that matches to a shelter address; *or* the word "homeless" is conveyed in the address field of the Be Counted Form and the city and/or county and state is provided.

All persons who meet the requirements for data capture will be counted in the Census 2000, irrespective of their records containing insufficient information for unduplication.

B. Input to Unduplication

Person records must contain the following information in order to be included in the unduplication process:

(1) First and last name (combined fields must contain at least 3 alphabetic characters)
(2) At least two 100% person characteristics, one of which is *date of birth* or *age*

C. Geographic Sort

Person records will be sorted and unduplicated within Local Census Office (LCO).

D. Person Matching

The person matching will be done by calling the matcher developed by Statistical Research Division. As part of the process, the individual matching characteristics (variables) will be assigned comparison weights based on their agreement or disagreement. Agreement weights are positive values and disagreement weights are negative values. Variables that are missing from one of the two person records involved in the comparison are assigned the weight of zero. A final weight assigned to the pair of person records is the sum of the agreement and disagreement weights for each matching characteristic. Cut off weights are applied outside the person match subroutine. A cut-off of 0.95 will be used for the Census 2000. The table below presents the variables and weights that will be used for the SBE unduplication.

Variable	Comparison	Agreement Weight	Disagreement Weight
First Name	string comparator	0.3864	- 0.2121
Last name	string comparator	0.3601	- 0.2184
Middle Initial	exact match	0.6896	- 0.6208
Sex	exact match	0.6994	- 0.6328
Race	exact match	0.5250	- 0.4750
Hispanic Origin	exact match	0.5250	- 0.4750
Month of Birth	exact match	0.4819	- 0.4274
Day of Birth	exact match	0.4819	- 0.4274
Year of Birth	range	0.5853	- 0.4513
Relationship	exact match	0.5013	- 0.4988

E. Person Selection

The person record is selected by applying a set of criteria to all of the person records until one is selected. The criteria for selection is:

1. Group Quarters (GQ) Type - select the person record by GQ type in the following sequence: (with long form taking priority over short form where applicable)

   a. 701-703
   b. 704-705

c. 706
d. BCF

2. Amount of data - select the person record with the most amount of short-form data.

3. For nonBCF record selection:

   a. Select the person record with the lowest GQ code.
   b. Select the person record with the lowest person identification number.

4. For BCF record selection, select the person record with the earliest data capture date.

C. SBE Unduplication Results

Flag SBE Unduplication Results as follows:

1=records with insufficient information for matching and are counted in Census 2000.
2=records that are not matched and are counted in Census 2000.
3=records that are matched to a Targeted Nonsheltered Outdoor Location (TNSOL) or BCF and are counted in Census 2000.
4=records that are not matched to a TNSOL or BCF and are counted in Census 2000.
5=records that are matched and not counted in Census 2000.

Prepare a file of all records that are input to the SBE unduplication. Refer to Attachment A "SBE Unduplication Results File" for the requirements of this file.

II BE COUNTED FORM GEOGRAPHIC ALLOCATION

The BCFs that do not match any person records in the SBE universe will be allocated to the SBE sites in the area by LCO using proportional allocation. These people are to be allocated ONLY to GQ types 701 and 704. If these specific GQ types are not in the LCO, then allocate the person record to *any* of the following GQ types within the LCO.

Preferred:
Allocate first to GQs which may yield similar characteristics (800s, 901, 902, 908, 400s).
If none of the above GQs are available, then allocate to the 100s.

Refer to Attachment B for a copy of the algorithm used to conduct the allocation.

Attachments

cc: DSSD Census 2000 Procedures & Operations Memorandum Distribution List
Service-Based Enumeration Team

SBE Unduplication Results File

A. Include the following fields from the **Decennial Master Address File:**

1. Geography

   a. State Code
   b. County code
   c. Type of Enumeration Area
   d. Local Census Office
   e. Assignment Area

2. GQ Type Code

   a. 701 — Shelters for People Without Conventional Housing
   b. 702 — Shelters for Children who are Runaways, Neglected, or Children Without Conventional Housing
   c. 703 — Shelters for Abused Women
   d. 704 — Soup Kitchens
   e. 705 — Regularly Scheduled Mobile Food Vans
   f. 706 — Targeted Non-sheltered Outdoor Locations

3. GQ ID Number

B. Include the following fields from the **Decennial Response File:**

1. Questionnaire Form Type (PFT)

   07 = D-10 "Be Counted"
   09 = D-15A "Individual Census Questionnaire" - short
   10 = D-15B "Individual Census Questionnaire" - long
   11 = D-20A "Individual Census Report" - short
   12 = D-20B "Individual Census Report" - long

2. Language of Data Capture Record (PCLANG)

   1 = English (US)
   2 = Spanish (US)
   3 = Chinese
   4 = Korean
   5 = Tagalog
   6 = Vietnamese
   7 = English (PR)
   8 = Spanish (PR)

3. Capture Date for BCF (IDC4)

4. GQ Person Sequence Number (PSEQ)

5. Person Name
   a. Last Name (PLNAME)
   b. First Name (PFNAME)
   c. Middle Initial (PMI)

6. Person Characteristics
   a. Sex (PSEX)
   b. Date of Birth (Month=PMOB, Day=PDOB, Year=PYOB)
   c. Age on April 1, 2000 (PAGE)
   d. Hispanic Origin (PSPAN01 - PSPAN05)
   e. Race Code (PRACE01-PRACE13)
   f. From the ICQ: Do you stay overnight MOST OF THE TIME at an apartment or house, at a shelter, on the street, or at some other type of place? (PONMOST)
   g. From the ICQ: Including today, how many days during the past 7 days did you receive a meal from a SOUP KITCHEN or MOBILE FOOD VAN? (PSOUPK)
   h. From the ICR and the ICQ: Including tonight, how many nights during the last 7 nights did you stay in a SHELTER? (PSNITES)
   i. From the ICQ: The Census Bureau has been counting people in this area. Did you fill out another Census form during the past week? (POTHFORM)

C. Add the following fields as a result of the **SBE Unduplication Process:**

Flag SBE Unduplication Results as follows:

1=records with insufficient information for matching and are counted in Census 2000.
2=records that are not matched and are counted in Census 2000.
3=records that are matched to a TNSOL or BCF and are counted in Census 2000.
4=records that are not matched to a TNSOL or BCF and are counted in Census 2000.
5=records that are matched and not counted in Census 2000.

D. Add the following fields as a result of the **Be Counted Form Geographic Allocation:**

1. Allocated GQ Type
2. Allocated GQ ID
3. Allocated geographic codes (DO, ARA, Block, Block part, and County code)

ATTACHMENT B

Be Counted Form Geographic Allocation

Persons enumerated on Be Counted Forms who marked the box indicating they have no address on April 1, 2000 will be proportionally allocated to shelters and soup kitchens for tabulation purposes.

1. Assign each SBE enumeration site a weight:

$$W_i = N_i / \sum_{i=1}^{S} N_i$$

where:

N_i = number of persons counted in the i-th SBE site (after unduplication); and,
S = number of sheltered and non-sheltered SBE sites in a LCO.
W = the relative weight of the facility within the LCO
(Use single precision to calculate W_i)

2. For each SBE_i i=1,2,...,S, define $L_i=L_{i-1} + W_i$ with $L_0=0$.

3. Generate a random number, RN_j, for each SBE Be Counted person.

   $0 \leq RN_j < 1$, j=1,2,...,P, where P = number of SBE Be Counted persons

4. Allocate the j-th SBE Be Counted person to the i-th SBE enumeration site for which:

   $L_{i-1} \leq RN_j < L_i$.

   Keep a count $P_i = P_i+1$ for each SBE_i of the number of Be Counted persons added.

5. The BCFs that do not match any person records in the SBE universe will be allocated to the SBE sites in the area by LCO using proportional allocation. These people are to be allocated ONLY to GQ types 701 and 704. If these specific GQ types are not in the LCO, then allocate the person record to *any* of the following GQ types within the LCO.

Preferred:
Allocate first to GQs which may yield similar characteristics (800s, 901, 902, 908, 400s).
If none of the above GQs are available, then allocate to the 100s.